Frequency-influenced choice of L2 sound realization and perception: evidence from two Chinese dialects

Yizhou Lan
Department of Asian and International Studies
City University of Hong Kong
eejoe.lan@gmail.com

Abstract

The study of second language speech perception usually put L1-L2 phonological mapping as the rule of thumb in predicting learning outcome, and seldom included more fine-grained aspects such as frequency. This study examines how frequency of sounds in L1 may influence L2 segmental production and perception, with examples from English learners native to two Chinese dialects, Cantonese and Sichuanese. Although these two dialects (L1s) have very similar phonological inventory, they produce certain L2 sounds in drastic difference. Productions of English voiceless interdental fricative and central liquid in the onset position were obtained in free speech from the two dialects' speakers in vast phonological environments. Then, perception tests, including AX and oddity tasks, were done for these two groups of speakers as well. Results showed that the two English sounds were respectively realized as different sounds in Cantonese and Sichuanese L1, which was reflected by both production and perception data. Findings suggest that L2 category formation is frequency-motivated instead of markedness-motivated, and is significantly influenced by the functional load of L1 sound input. Findings further imply that a quantitative and frequency-sensitive learning model is more suitable for L2 sound acquisition.

1 Introduction

Second language speech has generally seen as function of linguistic experience. However, how experience shape the formation of phonetic category was understudied. This study addresses a case when speakers from two L1s with similar segmental layout may have different realizations of L2 categories. Although theoretic models in speech learning such were very rich in literature, such as Perceptual Assimilation Model (PAM [1]) and its another version for L2 learners (PAM-L2 [2]) as well as Speech Learning model (SLM, [3]) had addressed different L1 assimilation patterns in learning multiple L2s, few studies had found similar multiple L1s yielding different L2 learning outcomes.

PAM and SLM suggest that second language learners will either assimilate the L2 sound categories (or sequence of sounds) to L1 sound categories according to different perceptual distances. Increased exposure to L2 will thus trigger distributive learning of L2 input by forming a new intermediate category between the L1 and L2 in the learner's common phonetic space [1]. In experience-based models, the positive effect of L2 exposure will increase the chance of distributive learning because the learnability of certain L2 categories should become stable if the input of L2 categories occurs in environments with similar frequency [3].

This paper displays that similar L1 inventories may result in different learning outcomes and argues that this phenomenon is influenced by frequency in similar ways as the native language was (NLM, [5]). The two English sounds under current investigation are the voiceless interdental fricative (/θ/) and the central liquid (/r/). In a pilot study, it was found that Sichuanse speakers replace English /θ/ by /s/ but Cantonese speakers by /f/.

28th Pacific Asia Conference on Language, Information and Computation pages 292–298

Also, Sichuanese speakers replace English /r/ by /z/ but Cantonese by /w/.

Previous literature has pointed out that these two English phonemes are difficult for Cantonese and Sichuanese learners to produce [6-8], but the question why the two dialects of Chinese may have different realizations of the sound was not addressed.

Cantonese and Sichuanese are both southern dialects of China. Cantonese and Sichuanese share a very similar consonant inventory in the onset position. Both dialects' onsets consist of bilabial, alveolar and velar plosives (/ph, th, kh, p, t, k/), as well as labiodental and alveolar fricatives (/f, s, z/). Nasals and liquids include /m, n, ŋ/. The only difference of the two dialects is that Cantonese does not have palatalized fricatives.

In the present study, Cantonese and Sichuanese L2 production and perception were examined. Firstly, the production of /θ/ was obtained from a sentence-making task, which contains stimuli words with /θ/. Then, the spectral envelope was analyzed through fast Fourier transformation (FFT) and sent to t-test for statistics [9]. For the production of /r/, same task was administered and the analysis was made into checking the F3 and waveform of /r/ (ibid.). Then, a perception test was designed. Native speakers' production was presented to another two groups of speakers and they were required to identify from two sounds and discriminate from three sounds, which were cross-checked with the production indications. For example, both Cantonese and Sichuanese speakers listened to /f/ and /s/ tokens against /θ/ in a task, and /w/ and /z/ in another one.

2 Method

2.1 Participants

Effort was made to control all the biographical, affective and experiential factors of the two groups of participants. 8 Cantonese and 8 Sichuanese speakers, with equal numbers of males and females, were recruited. Both groups of speakers were experienced learners of English, with the age of acquisition of English (AOA) earlier than 7 years old. A group of native speakers of the Standard American English also participated in the study.

Cantonese speakers were not exposed to formal instruction of any other languages, and their parents speak any other languages other than Cantonese (including English). The situation for Sichuanese speakers is more complex. Since speaking Mandarin at school is mandatory, and those with early English AOA have all attended school, they have been exposed to Mandarin as well as Sichuanese. This has brought about a difference of these two groups of speakers. However, it cannot be eliminated due to language policy [10].

2.2 Stimuli and Procedure

We designed a production and a perception test to find out whether L2 category formation (/θ/ and /r/) is different for Cantonese and Sichuanse speakers; and we retrieved the functional load of these sounds on a small-scale corpus to see if frequency motivates the difference of categorical formation.

For the production experiment, stimuli contained experiment words (/r/ with 5 vowels and 3 syllable structures; /θ/ with 5 vowels and 4 syllable structures, with ten repetitions respectively: e.g., rit, ree, rin; θit, θee, θin) control words (/f/ /s/ /w/ /z/ with 5 vowels and 3 syllable structures, with ten repetitions; e.g., *fit, sat, wut, zot*) and filler words with other onsets (/p/, /t/, /k/ as the same structures, with five repetitions).

The experiment procedure was a semi-free speech with given stimuli. Participants were asked to make five stories with the given words, each story containing two sentences. The words were later cut out of the sentence for analysis. Most of the stimuli words were obtained after a long pause at the intonational phrase level so that phonetic environment will not influence too much of the production. For the /θ/ contrast, the spectral energy concentration was analyzed for the characterization of /s/ or /f/ contrast (here, some productions were too short and taken as /t/ tokens). Participants were not aware of the purpose of the study. They were informed that they were participating in a test testing fluency in spoken English.

As we aim to dig out the characteristics of actual vernacular form of speech instead of citation forms, we did not strictly control the number and order of occurrence of stimuli, but still controlled phonetic environment and the number of tokens. Altogether 101 usable tokens (including /s, f, θ, r, w, z/-initials) were collected from 8 Cantonese and 8 Sichuanese student participants' productions and 48 tokens from the native English participant's productions (101+48=149 tokens). The productions were cut out of the sentence and segmented as phonemes within those words. The onset parts of the productions, defined as the section from the beginning of waveform to the steady state of vowel, were examined for in spectral analysis.

The perception study was done in the same laboratory. Both an AX task and an oddity task (a variation of the ABX task) were performed. In the AX task, listeners were presented with two stimuli and they need to identify it is either /θ/ or /f/ or /s/. In the oddity task, they were given three stimuli in ABA, ABB or AAB form to distinguish. They need to decide which one is different. Theoretically the token number to be included in analysis was 27 stimuli × 5 repetitions × 2 combinations + 27 stimuli × 5 repetitions × 3 combinations = 675 tokens for each speaker. After screening, a total of 620 tokens were selected as the perception test material. Within-trial inter-stimuli interval (ISI) was set at 50ms and between-trial ISI at 200ms. All trials were randomized and added with equal numbers of fillers.

Since the relationship of frequency and category assimilation patterns was to be investigated, the third step of the current study was the extraction and comparison of functional load data from a corpus of two dialects and relating of the functional load to the empirical study (including production and production) results. Since Sichuanese does not have an established corpus to date, we used the entries of a published wordlist and annotated them with productions in Cantonese and Sichuanese, which controlled the word frequencies in these two dialects. The choices of words from *Xiandai Hanyu Changyong Zibiao* [11], a list of 2500 most commonly used Chinese characters to relate the phonological families of Chinese dialects. Word frequency was considered as a coefficient of the calculation of sound frequency count. We then examined the correlation between the assimilation pattern and functional load. It was a limitation not being able to employ more cognitive methods to establish a causal link between the two instead of a weak, correlational one, but due to technical reasons, the attempt was not realized.

3 Results

3.1 Production test

Spectral envelopes of the fricative productions were analyzed for Cantonese and Sichuanese speakers. First, the /f/ and /θ/ sounds were compared for similarity for both Cantonese and /s/ and /θ/ for Sichuanse speakers. For the /z/ and /r/ contrast, since these two sounds are easy to distinguish, sound with formant will be classified as /r/.

As the study aims not to find the criteria of identifying the fricatives but distinguishing them in shape, we are focusing on the peak of energy concentration instead of spectral moments. The average peak for Cantonese production of /f/, /s/ and /θ/ were 6754, 7259 and 6145 respectively for Cantonese speakers. For Sichuanese speakers, the figures were 6248, 7195 and 7246. Between-group variance tests show that the difference was insignificant for spectral peak. However, within the Cantonese speakers, the difference is significant for /s/ and /θ/ [$F(2, 248)=3.488$, $p<.0001$] not /f/ and /θ/ showed by an ANOVA test. The Sichuanese data was reversed, i.e. significant for /f/ [$F(2, 248)=2.125$, $p<.001$] but not for /s/. The results indicate that Cantonese speakers' production of /θ/ was similar to /f/ but different from /s/, and for Sichuanese, vice versa (see Figure 1 for an example of the Cantonese case. The energy concentrations of /θ/ overlap significantly more on /f/ than /s/).

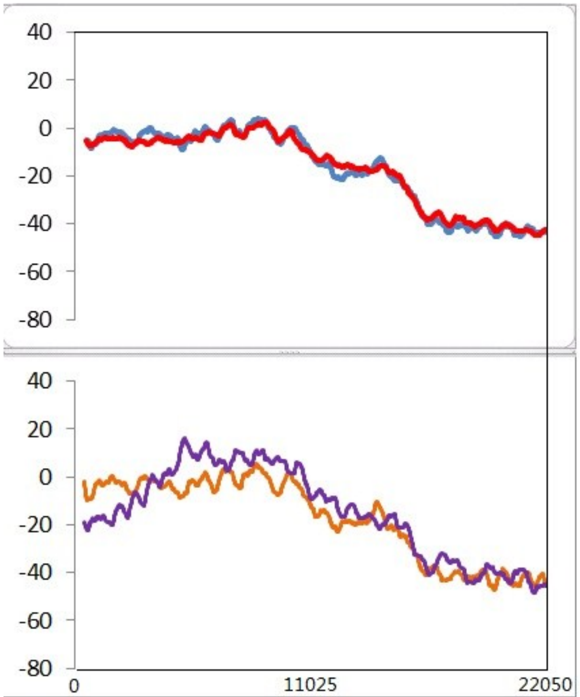

Figure 1: Comparison of sound pressure for /f/ and / θ / (upper) and /s/ and / θ / (lower) central spectrum. Measurement was done with 50ms pre-emphasis. The y-axis is in dB and y-axis in Hz.

The average duration for /s/, /f/ and /θ/ were 55, 65 and 47 ms respectively by Cantonese speaker, 53, 80 and 45 ms by Sichuanese speakers. The difference is not significant (see Figure 2).

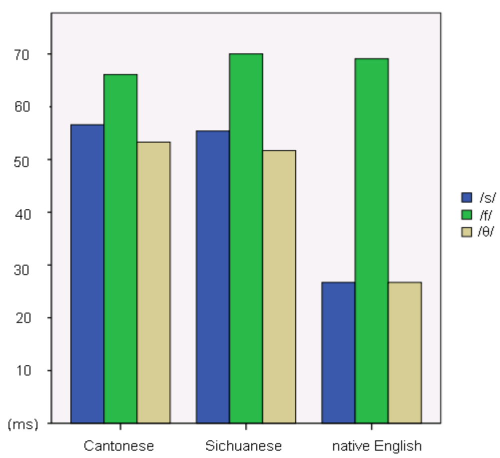

Figure 2: Duration of frication of /s, f, and θ/ by Cantonese, Sichuanese and English speakers.

However, for English speakers, the /θ/ and /f/ duration was much smaller, as 32 and 33 ms. As confirmed by previous studies (Flege and Wang, 1996), Chinese speakers of English did not distinguish fricative duration as native English speakers did, probably due to the syllable timing. Within-group variance tests shows that the difference was insignificant but significant for duration comparing between Cantonese and Sichuanese groups, [F(2, 248)=1.154, p=.248] but near-significant within groups [Cantonese: F(2, 124)=2.459, p=.065; Sichuanese F(2, 124)=3.245, p=.071] (see Figure 2).

For the /r/ contrast, the spectrogram of both Cantonese and Sichuanese speakers was examined. Formant contours and affrication was analyzed qualitatively. Only Sichuanese productions were seen of affrication indicating the presence of /z/, whereas Cantonese speech showed considerable F2 and F3 changes which could be seen as intermediate instances between /r/ and /w/. From above production data, reversed production patterns were shown for both /f/ and /s/ for /θ/ as well as /w/ and /z/ for /r/.

3.2 Perception test

Overall speaking, the identification and discrimination test result showed that the perceptual accuracy was 61.3% by the five Cantonese speakers, and 66.7% by Sichuanese speakers. For Cantonese speakers, the difference on [F(3, 617)=8.719, p<.0001], but not for English speakers [F(3, 617)=1.249, p=.576]. The effect of task was not significant. Due to such insignificance, identification and discrimination task results were computed into average and represented as /x/-/y/ accuracy rates for the ease of comparison.

For Cantonese speakers, vowel differences were not significant. Accuracy rate for /θ/ and /s/ discrimination was 85.75%, and accuracy rate for /θ/ and /f/ was 56.5%. Such a difference was significant [t=2.128, df=317, p<.0001]. Accuracy rate for /r/ and /w/ was 88.15%, /r/ and /z/ was 71.25%. The difference was near-significant [t=-0.257, df=317, p=.042].

For Sichuanese speakers, vowel differences were not significant as well. Accuracy rate for /θ/ and /s/ discrimination was 42.15%, and accuracy rate for /θ/ and /f/ was 82.45%. Such a difference was significant [t=2.719, df=317, p<.0001]. Accuracy rate for /r/ and /w/ was 67.5%, /r/ and /z/ was 78.85%. The difference was not significant [t=5.124, df=317, p<.0001] (See Figure 3).

As a random factor, individual difference within both groups did not significantly influence the perceptual accuracy.

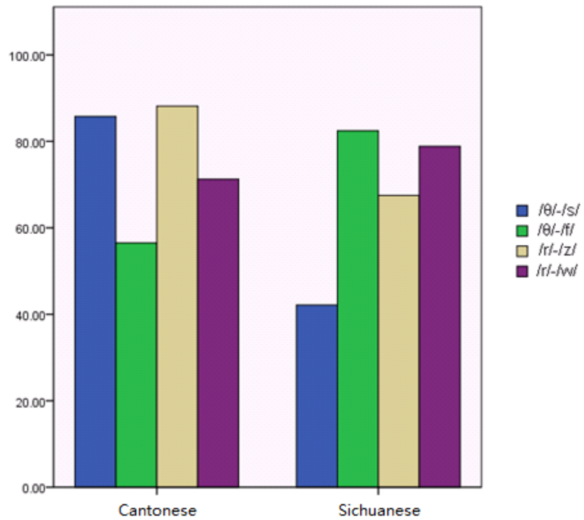

Figure 3: Comparison of mean perceptual accuracy rates of Cantonese and Sichuanese speakers.

3.3 Comparison of Frequency

The following table layouts the item under discussion, and dominantly assimilated sound as acquired from 3.1 and 3.2. For example, the dominant choice of realization and perception for Cantonese /θ/ was /f/ instead of /s/.

To investigate whether frequency was parallel to the assimilation patterns, the functional load of the two word-lists in Cantonese and Sichuanese was compared. The result summarized from the above experiment was shown in Table 1.

Item	Dominant	Item	Dominant
C /s/	/f/	C /r/	none
S /s/	/s/	S /r/	none
C /f/	/f/, /h/	C /z/	/z/
S /f/	/f/	S /z/	/z/, /r/

Table 1: Dominant sound category in Cantonese (C) and Sichuanese (S) speech.

According to its definition, functional load (FL) of two contrasting sounds is calculated as the function of frequency of a lexical entry and the frequency of the two involving sounds, which can be expressed as follows in (1):

$$FL(x,y) = \frac{H(L) - H(L_{xy})}{H(L)} \qquad (1)$$

A report showed that in American English, the functional load of /f/ and /θ/ was 1×10^{-3}, and 2×10^{-3} for /s/ and /θ/ [12]. Therefore we could see that for English, the sound /s/ is actually more frequently confused with /θ/ than /f/, and the choice by Cantonese speakers may be not reflecting the English L1 predictions. Here we could see that the functional loads for fricatives are different across the two dialects of Chinese. The functional load calculated for Cantonese and Sichuanese /s, f/ pair and /z, w/ pair was displayed in Table 2.

Sound pair	Functional load in most used Chinese characters
Cantonese /s vs. f/	0.125
Sichuanese /s vs. f/	0.750
Cantonese /z vs. w/	0.054
Sichuanese /z vs. w/	0.375

Table 2: Functional load in onset position in 2500 most used Chinese characters.

From the data, we could see that /f/ is functionally more loaded than /s/ for Cantonese speakers, and vice versa for Sichuanese speakers. On the contrary, /w/ was more functionally loaded for Sichuanese than for Cantonese.

4 General Discussion

Production results showed that the role of functional load did differ in Cantonese and Sichuanese, and the more frequent and more functionally loaded /f/ in Cantonese, compared with Sichuanese, was linked with the choice of /f/ rather than /s/ in the realization and perception of /θ/. Conversely, the Sichuanese choice also preferred the more functionally loaded one, /s/. The same patterned preference showed for /w/ and /z/ in Cantonese and Sichuanese as well.

The spectral differences in Cantonese and Sichuanese L2 English lied in spectral envelope, esp. spectral peak. However, patterns of duration of the fricatives were not significantly different amongst these two groups of speakers, maybe because this dimension of acoustic information was not distinguished by both Cantonese and Sichuanese speakers as a whole [13].

Perceptually, since it is clearly shown that the value of accuracy rates was reversed for Cantonese and Sichuanese speakers, and the inclination was especially true for the /θ/ sound. Therefore it could be drawn from the results that Cantonese and Sichuanese speakers of English have different perception of sound categories, and apply different assimilation routes to sounds.

For both Cantonese and Sichuanese learners, perception and production of these sounds were quite symmetric. It further suggested a steady tendency of difference in choice of L2 realizations for these two dialects, though their inventories were of very similar layout.

Despite the production and perception results which showed a different inclination towards /f/ and /s/ by Cantonese and Mandarin learners, such conclusion is apt to test by a question whether the difference is due to phonetic closeness as proposed by J. Jenkins. However, the design of the study confirmed that the phonetic distance of /s/-/θ/ and /f/-/θ/ acoustically is similar.

The current results shed light on the crystallization of two significant theoretic debates. The first debate involves whether L2 speech realizations are mapped on discrete phonological units, i.e., phonemes, or through more distributional processes which is influenced by the frequency of the L1. The first approach, including Optimality Theory, cannot explain the data in the current study because although OT is based on gradable constraints, it still believes that the output is the same for similar L1 phonological structures. More importantly, the difference in outputs for these two dialects is not markedness-motivated but frequency-motivated. The OT claim of tearing linguistic performance into perceptual level and representational level [14] is more complex than this frequency-based explanation.

The second debate which concerns this study is the choice of assimilation routes by L1 only or by a cluster of dynamic frequency correlates of L1 (and maybe experiences on other languages). In SLM's suggestion, assimilation is based on perceived acoustic similarity only, but the results here showed that an assimilation route can be dynamic and may be influenced by the functional load of L1. This probabilistic view is in line with the basic assumption of the NLM model [5] but slightly different from SLM in that it opposes discrete assimilation pattern projections from distance to learning outcome. A probabilistic model predicts assimilation outcomes not based on the distance, but on the instances on the input of L2 phones, and its probabilistic balance with regard to L2. In other words, L2 learning is statistical learning instead of a mere calculation of distances.

Although native English speakers perceive /s/ as a better exemplar of /θ/ compared with /f/ due to the higher functional load, Cantonese speakers prefer /f/ in a very clear-cut manner. It is implied that L1 frequency is such an important factor that can override L2 preferences, which also exists in the input in their learning. L1, in the frequency's perspective, plays a more important role than L2 even after many years of learning. This phenomenon also challenges learnability of some L2 categories, since according to SLM, the categories should receive even more influence on L1 and L2 input and establish an intermediate category provided exposure to the L2. However, as the result suggests, the preference of /f/ by Cantonese speakers cannot be eliminated and thus cannot be learned in a small time span.

Findings indicate that the mechanism for L2 categorical formation is more than a perception-production chain, and may involve statistical learning effects. When the prediction through phonological categorical assimilation and frequency-based predictions collide, the latter is favored. However, there might also be other variables stretching outside the realm of phonetics and phonology that influence the results, because the affective factors of this study were not fully controlled. Future studies should involve more specific measurements to mine out these variables.

.

Acknowledgments

The author thanks Karen Chow for data collection, and Weiping Zhong for providing word frequency measurement.

References

[1] C. T. Best, "A direct-realist view of cross-language speech perception," in Strange W [Ed.] Speech

perception and linguistic experience: Issues in cross-language research, 171–204, 1995.

[2] C. T. Best, and M. D. Tyler, “Nonnative and second-language speech perception: Commonalities and complementarities”. In Munro M. J. & Bohn O.-S. [Eds.] Second language speech learning: The role of language experience in speech perception and production, 13-34, 2007.

[3] J. E. Flege, “Assessing constraints on second-language segmental production and perception,” Phonetics and phonology in language comprehension and production: Differences and similarities, 319-355, 2003.

[4] B. Hayes, and C. Wilson, “A maximum entropy model of phonotactics and phonotactic learning,” Linguistic Inquiry, 39(3), 379-440. 2008.

[5] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, “Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e),” Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1493), 979-1000, 2008.

[6] D. Rau, Chang, H. H. A., and Tarone, E. E. “Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative”. Language Learning, 59(3), 581-621, 2009.

[7] Hung, T. N. “Towards a phonology of Hong Kong English”. Bolton, K. ed. Hong Kong English: autonomy and creativity (Vol. 1),119-140, 2002.

[8] Lee, P. W. “The study of English in China with particular reference to accent and vocabulary”. Master Thesis submitted to The university of Hong Kong, 2002.

[9] R. D. Kent, and C. Read, “The acoustic analysis of speech,” Thomson Learning Albany, NY, 2002.

[10] S. Evans, “The Long March to Biliteracy and Trilingualism: Language Policy in Hong Kong Education Since the Handover,” Annual Review of Applied Linguistics, 33, 302-324, 2013.

[11] State Language and Letters Committee of China, “List of Commonly Used Characters in Modern Chinese, *Xiandai Hanyu Changyong Zibiao*”, 1988.

[12] S. Dinoj, and N. Partha, “Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals”, Nedergaard Thomsen, Ole [Ed.] Current trends in the theory of linguistic change. In commemoration of Eugenio Coseriu (1921-2002). Amsterdam & Philadelphia: Benjamins, 2006.

[13] W. Strange, “Automatic selective perception (ASP) of first and second language speech: A working model,” Journal of Phonetics, 39(4): 456–466, 2011.

[14] J. Dekkers van der Leeuw, J. M. van de Weijer, “Optimality Theory: Phonology, Syntax, and Acquisition,” OUP, 2000.